

# SURROGATE DATA FOR NON-STATIONARY SIGNALS

ANDREAS SCHMITZ AND THOMAS SCHREIBER

*Physics Department, University of Wuppertal,  
D-42097 Wuppertal, Germany*

Standard tests for nonlinearity reject the null hypothesis of a Gaussian linear process whenever the data is non-stationary. Thus, they are not appropriate to distinguish non-linearity from non-stationarity. We address the problem of generating proper surrogate data corresponding to the null hypothesis of an ARMA process with slowly varying coefficients.

## 1 Introduction

Most methods used in the field of linear and nonlinear time series analysis assume stationarity of the considered data. Non-stationarity is very likely to lead to wrong results. This is especially true for tests for nonlinearity. A common approach is to split the time series into segments which can be considered nearly stationary and perform individual tests. But for short time series or not too slowly varying non-stationarities these segments have to be made too short to meaningfully calculate a test statistic on them.

## 2 Surrogate Data

To generate surrogate data that has the same time variation of autocorrelations than the original data one can in principle follow the general approach of constrained randomization.<sup>1</sup> The corresponding cost function would be the discrepancy between the autocorrelations of the original and the surrogate data in sliding windows. Autocorrelations should be calculated up to sufficiently high lags. The most striking disadvantage of this method is the extreme computational burden. Therefore, we use an alternative method:

1. split up the time series into segments
2. generate a surrogate for each segment
3. join the segments to one new surrogate

For step 2 we use an iterative algorithm.<sup>2,6</sup> The size of the segments is typically too small to perform individual tests for nonlinearity on them, because most test statistics need sufficiently large data sets. A disadvantage of the method is the loss of correlations *between* the segments which can lead to a bias.

## 3 Cyclostationary Processes

The first class of processes we consider are cyclostationary, having periodically varying parameters. These processes have been found to yield rejections of the null hypothesis<sup>3</sup> when using a test statistic derived from the correlation sum. Here we

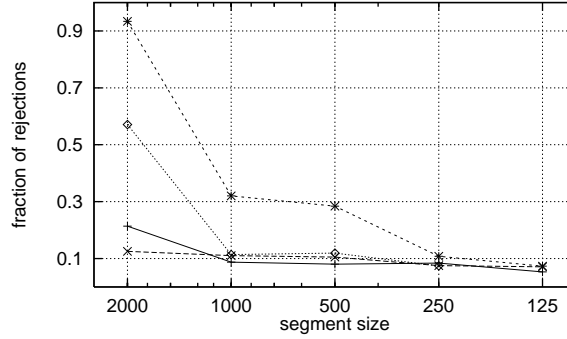


Figure 1. Size of nonlinearity tests with varying segment sizes. The modulation period of the non-stationarity is fixed to  $T_{\text{mod}} = 2000$ . From upper to lower curve the values  $A = 2.0, 1.5, 1.0$  and  $0.0$  are used.

use the maximum likelihood estimator of the correlation dimension<sup>4,5,6</sup>

$$D_2^{\text{ML}} = \frac{C(\epsilon_0)}{\int_0^{\epsilon_0} C(\epsilon)/\epsilon d\epsilon}. \quad (1)$$

Following Timmer<sup>3</sup>, the considered time series is generated by an AR(2)-process

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \eta_n, \quad n = 1, \dots, 2000 \quad (2)$$

with coefficients chosen to be

$$\begin{aligned} a_1 &= 2 \cos(2\pi/T) \exp(-1/\tau) \\ a_2 &= -\exp(-2/\tau) \end{aligned} \quad (3)$$

where  $T$  is the period and  $\tau$  the relaxation time of a damped oscillator. We will use  $T = 7$  and  $\tau = 50$ . Non-stationarity is introduced by a modulation of the period  $T$

$$T(n) = T_{\text{mean}} + A \sin(2\pi/T_{\text{mod}} n). \quad (4)$$

By varying  $A$  and  $T_{\text{mod}}$  the degree and time scale of the non-stationarity can be changed which is reflected in non-constant running variances and autocorrelations.

### 3.1 Testing the test size

The fraction of tests that falsely reject the null hypothesis, when it is in fact true, is called the *size* of the test. Since a single test<sup>3</sup> says not much about the size, we performed 1000 one-sided tests with 9 surrogates for each process. A correct test with proper surrogates should yield the nominal size  $\alpha = 0.1$ .

For the first tests we used  $T_{\text{mod}} = 2000$  for the modulation period. As shown in Fig. 1 the new method does indeed yield the correct size of the test for a segment size of about  $T_{\text{mod}}/2 = 1000$ . Only for relatively strong non-stationarity ( $A = 2.0$ ) smaller segments seem to be necessary. If we reduce the time scale of the non-stationarity by setting  $T_{\text{mod}} = 1000$  we get sizes shown in Fig. 2. As expected,

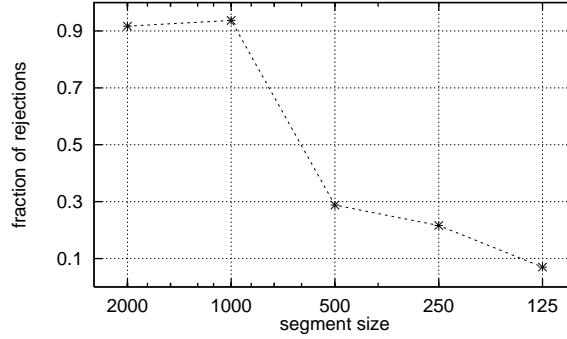


Figure 2. Size of nonlinearity tests with varying segment sizes. Now  $T_{\text{mod}} = 1000$  and  $A = 2.0$  are used.

only segment sizes which are smaller than  $T_{\text{mod}}$  can decrease the number of false rejections.

#### 4 AR(1)-Processes

We found cases in which non-stationarity does not yield a higher rate of rejections of the null hypothesis but just the opposite. Consider an AR(1)-process

$$x_n = a_1 x_{n-1} + \eta_n \quad (5)$$

with a nonlinear measurement function  $M(\cdot)$

$$M(x_n) = x_n \sqrt{|x_n|}. \quad (6)$$

Non-stationarity is introduced by varying  $a_1$  from 0.5 to 0.9 along the time series. We now choose a *prediction error* as test statistic.<sup>6</sup> It is based on locally constant predictions in an  $m$ -dimensional embedding space.

Again we performed 1000 one-sided tests with 9 surrogates to estimate the size. Additionally we calculated the rank of the original data within the surrogates going from 1 if the prediction error of the original data set was lower than for all surrogates up to 10 if the original data yields the largest prediction error. For usual stationary surrogates we get the distribution presented on the left side of Fig. 3. In 35% of the tests the original data has a *larger* prediction error than all the surrogates and the null is rejected in only 2.3% of the tests. This is the opposite of what we expect for non-linear signals, which should be better predictable. Here, predictions are easier for the stationary surrogates with constant autocorrelations than for the original non-stationary data. Prediction error is not a useful discriminating statistic in this case.

On the right side of Fig. 3 the results for tests with the new surrogates and different segment sizes are shown. For a segment size of 1000 we get the correct size and therefore a uniform distribution. With decreasing segment sizes the prediction error of the surrogates gets larger than for the original data. The predictions seem to be very sensitive to the autocorrelations that get lost between the segments.

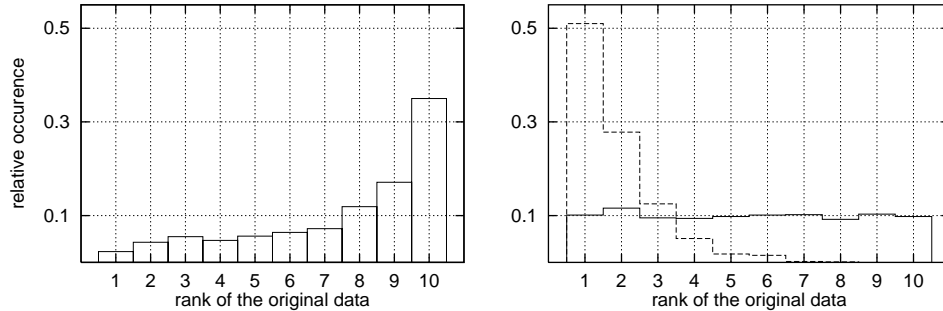


Figure 3. Distributions of the rank of the original data within the surrogates for stationary surrogates (left) and surrogates with segment sizes 1000 (solid) and 250 (dashed).

## 5 IV. Summary

Dealing with non-stationary signals is still a difficult business. We have given numerical evidence that in a test for nonlinearity of a data set with possible non-stationarity in form of slow variation it may be feasible to generate surrogates on separate segments. These segments can be made smaller than would be necessary for individual nonlinearity tests. But the segment size and the test statistic should still be chosen with great care, since no general theory guarantees the correctness of the test. As a final remark we want to stress that the reasoning in this work is not applicable to sudden changes like jumps or spikes.

## References

1. T. Schreiber, *Constrained randomization of time series data*, Phys. Rev. Lett. **80**, 2105 (1998).
2. T. Schreiber and A. Schmitz, *Improved surrogate data for nonlinearity tests*, Phys. Rev. Lett. **77**, 635 (1996)
3. J. Timmer, *The power of surrogate data testing with respect to non-stationarity*, Phys. Rev. E **58**, 5153 (1998)
4. F. Takens, *On the numerical determination of the dimension of an attractor*, in: B. L. J. Braaksma, H. W. Broer und F. Takens, eds., *Dynamical Systems and Bifurcations*, Lecture Notes in Math. Vol. 1125, Springer, Heidelberg-New York (1985)
5. J. Theiler, *Lacunarity in a best estimator of fractal dimension*, Phys. Lett. A **135**, 195 (1988)
6. R. Hegger, H. Kantz, and T. Schreiber, *Practical implementation of nonlinear time series methods: The TISEAN package*, accepted by CHAOS (1999). The software package is publicly available at <http://www.mpi-pks-dresden.mpg.de/~tisean>.